# Stochastic Processes
# Final Exam Review Notes
# (Fall 2019)

Maxie D. Schmidt

Revised: December 8, 2019

---

# 1 SUMS OF POISSON AND EXPONENTIAL RV

## 1.1 Poisson random variables

**Definition 1.1** (Poisson Distribution)**.** The *Poisson distribution* is defined by a single rate parameter $\lambda > 0$. It's discrete pmf, $f : \mathbb{N}_0 \to \mathbb{R}$, is defined by $f(n) := \frac{\lambda^n}{n!} e^{-\lambda}$ for $n \geq 0$. We denote a random variable with this distribution by $\xi \sim \text{Poisson}(\lambda)$. Consequently, $\mathbb{E}[\xi] = \text{Var}[\xi] = \lambda$.

**Proposition 1.2** (Sums of IID Poisson Random Variables)**.** *Let $X_1, \ldots, X_k$ be iid $X_i \sim \text{Poisson}(\lambda_i)$. Then* $\sum_{i=1}^{k} X_i \sim \text{Poisson}\left(\sum_{i=1}^{k} \lambda_i\right)$.

**Proposition 1.3** (Poisson Distributions as a Limit)**.** *Let $\lambda$ be fixed. For any $m \geq \lambda$, let $b_1^{(m)}, \ldots, b_m^{(m)}$ be iid* $\text{Bernoulli}(\lambda/m)$*, and let the $f_m$ denote the pmf of $\sum_{i=1}^{m} b_i^{(m)}$. Then the sequence $(f_m)_{m \geq \lambda}$ converges pointwise to the pmf of a random variable with distribution* $\text{Poisson}(\lambda)$.

## 1.2 Exponential random varaibles

**Definition 1.4** (The Exponential Distribution)**.** Let $\lambda > 0$ be fixed. Define $F(x) := 1 - e^{-\lambda x}$ to be the CDF of an *exponential random variable* with rate $\lambda$. The PDF of such a random variable is given by $f(x) = \lambda e^{-\lambda x}$ for $x > 0$. We have that if $\rho \sim \text{Exp}(\lambda)$, then $\mathbb{E}[\rho] = \frac{1}{\lambda}$.

**Proposition 1.5** (Sums of IID Exponential Random Variables)**.** *Let $Z_1, \ldots, Z_m \sim \text{Exp}(\lambda)$. Then the CDF of the sum is given by*

$$\mathbb{P}[Z_1 + \cdots + Z_m \leq t] = 1 - w^{-\lambda t} \sum_{k=1}^{m-1} \frac{(\lambda t)^k}{k!}.$$

# 2 THE METROPOLIS-HASTINGS ALGORITHM

## 2.1 Setup, configuration, and initial inputs

Suppose that $S$ is a finite state space and that $(X_n)_{n\geq 0}$ is an irreducible Markov chain with stationary distribution $\pi$. The goal of the algorithmic procedure discussed next is to modify the existing chain so that it has a new stationary distribution $\pi'$, which we will construct and define precisely below.

## 2.2 Definitions

Let $P$ denote the transition matrix of the original chain. For any fixed $i, j \in S$, we define the *acceptance ratio* to be

$$A(i,j) := \min\left(1, \frac{\pi'(i)}{\pi'(j)} \cdot \frac{p_{ij}}{p_{ji}}\right),$$

where $p_{ij}$ denotes the $(i,j)^{th}$ entry of $P$. We then define the transition matrix, $P'$, of the modified chain according to the rules

$$p'_{ij} := \begin{cases} p_{ij}A(j,i), & j \neq i; \\ 1 - \sum_{j\neq i} p_{ij}A(j,i), & i = j. \end{cases}$$

We can verify that the stationary distribution of the new, modified chain is actually $\pi'$ (with any $i \in S$):

$$\begin{aligned} \sum_{j=1}^{N} p'_{ij}\pi'(j) &= \sum_{j\neq i} p_{ij}A(i,j)\pi'(j) + \left[1 - \sum_{j\neq i} p_{ij}A(j,i)\right]\pi'(i) \\ &= \pi(i) + \sum_{j\neq i} \min(p_{ji}\pi'(j), p_{ij}\pi'(i)) - \sum_{j\neq i} \min(p_{ij}\pi'(i), p_{ji}\pi'(j)) \\ &= \pi'(i), \forall i \in S. \end{aligned}$$

Since this relation holds for all $i \in S$, the above proof implies that $\pi'$ is the stationary distribution of the modified chain.

### 2.2.1 Consequences

If the original chain is irreducible, then the modified chain is also irreducible.

## 2.3 Application to the simple random walk on a connected undirected graph

We use the algorithm defined above to construct a uniform stationary distribution of the vertices of a connected, undirected graph. Let $G$ be a graph on the vertex set $V := \{1, 2, \ldots, N\}$, and suppose that $(X_n)_{n\geq 0}$ is a simple random walk on $G$ with corresponding stationary distribution $\pi$. We recall from previous lectures that

$$\pi(i) = \frac{\deg(i)}{\sum_{j\in V} \deg(j)}.$$

Notice that unless the graph is regular, the stationary distribution $\pi$ is *NOT* uniform on the vertex set!
**Goal:** Update the chain using the algorithm to make the resulting stationary distribution uniform.
**Modifications:** For this application, we have the following definition of the acceptance ratio:

$$A(i,j) = \begin{cases} \min\left(1, \frac{\deg(j)}{\deg(i)}\right), & i \in N(j); \\ 1, & \text{otherwise}. \end{cases}.$$

It follows that the transition matrix $P'$ defined by this formula yields the modified chain with a uniform stationary distribution.

**Proposition 2.1.** *Let $(X_m)_{m\geq 0}$ be an irredicuble, aperiodic Markov chain on $S := \{1, 2, \ldots, N\}$. Then*

$$\lim_{m\to\infty} \|\mathbb{P}[X_m \in \cdot] - \pi\|_{TV} = 0.$$

# 3 RENEWAL REWARD PROCESS

## 3.1 Definition of this type of process

Suppose that $y_i$ are iid random variables with *interarrival distribution* given by $F$. Then we define the *renewal sequence* $(S_n)_{n\geq 0}$ by

$$S_n := \sum_{i=1}^{n} y_i.$$

If $S_0 := 0$, the we call the sequence *pure*. Otherwise it is called *delayed*. The $S_n$ are called *renewal times*, or *renewal epochs*. For real $t > 0$, we define the *counting function* to be

$$N(t) := \#\{n : S_n \leq t\}.$$

## 3.2 The law of large numbers for this process (with proof)

**Theorem 3.1** (SLLN for Renewal Sequences)**.** *Suppose that $\mu := \mathbb{E}[y_i] < \infty$. Then a.e.,*

$$\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\mu}.$$

*Proof (Sketch).* We sketch the key component steps in the proof of this SLLN variant below:

- By the (ordinary) SLLN,
$$\frac{1}{n}\sum_{i=1}^{n} y_i \xrightarrow[n\to\infty]{a.e.} \mu.$$
- We observe that $N(t) \to \infty$ as $t \to \infty$.
- By combining the first two steps, we have that
$$\frac{1}{N(t)}\sum_{i=1}^{N(t)} y_i \to u \iff \lim_{t>0, t\to\infty} \frac{S_{N(t)}}{N(t)} = u, \text{a.e.}$$
- Finally, $S_{N(t)-1} \leq t \leq S_{N(t)}$.

In conclusion, we obtain that a.e.,

$$\frac{S_{N(t)-1}}{N(t)-1} \cdot \frac{N(t)-1}{N(t)} \leq \frac{t}{N(t)} \leq \frac{S_{N(t)}}{N(t)} \implies \lim_{t\to\infty} \frac{t}{N(t)} = \mu. \qquad \square$$

**Theorem 3.2** (CLT for Counting Functions)**.** *Suppose that $(S_n)$ is a renewal sequence, that $\mathbb{E}[y_i] := \mu$, and* $\mathrm{Var}(y_i) := \sigma^2 < \infty$. *Then as $t \to \infty$, the distribution of the random variable $\frac{N(t)-t\mu^{-1}}{(t\sigma^2\mu^{-3})^{1/2}}$ converges to the CDF $\Phi$ of the standard normal distribution. Moreover,*

$$\frac{\mathbb{E}[N(t)]}{t} \xrightarrow{t\to\infty} \mu^{-1}.$$

## 3.3 Relation to point processes

**Definition 3.3.** A *point process* on $\mathbb{R}$ is a random disrete measure on $\mathbb{R}$ of the form $\xi = \sum_{m=0}^{U} \delta_{X_m}$ where $U, (X_m)_{m\geq 0}$ are random variable and $U$ is allowed to be infinite. Such a process assigns each fixed Boral subset $I \subset \mathbb{R}$ a random number denoted by $\xi(I) = |\{m : X_m \in I\}|$.

We have the following fact: For every renewal process on $\mathbb{R}^+$, there is an associated point process of the form $\xi = \sum_{n\geq 0} \delta_{S_n}$.

## 3.4 Example: Operation interval of a machine

The operation of a machine is uniformly distributed on the integers $\{0, 1, \ldots, 50\}$ (in days). After it breaks down, the machine is repaired for 3 days, and then the cycle repeats.

1. What is the long-term percentage of inoperable days?
2. Estimate the probability that within a given year there are at least 42 inoperable days.

(Solution on page 44 of notes.)

## 3.5 Relation to the renewal equation

See the full introduction to the renewal equation starting on page 13 first. Suppose we have a pure renewal sequence $(S_n)_{n\geq 0}$, $S_0 = 0$, $S_n := \sum_{i=1}^{n} Y_i$ for $n \geq 1$, where $(Y_i)_{i\geq 1}$ is iid. The ordinary counting function $N(t) := |\{n \geq 0 : S_n \leq t\}|, t \geq 0$ is defined on an interval as $N[a, b] := |\{n : S_n \in [a, b]\}|$. We can compute the expected values in terms of $F(t)$, the CDF of $Y_1$ and where $F^{\odot 0} := \mathbb{1}_{[0,\infty)}$:

$$\begin{aligned}
\mathbb{E}[N(t)] &= \sum_{n\geq 0} \mathbb{P}[S_n \leq t] = \sum_{n\geq 0} F^{\odot n}(t) \\
&= F^{\odot 0}(t) + F \odot \left( \sum_{n\geq 1} F^{\odot (n-1)} \right)(t) \\
&= F^{\odot 0}(t) + \left(F \odot \mathbb{E}[N(\cdot)]\right)(t).
\end{aligned}$$

Thus we see the general form of the renewal equation is given by

$$Z(t) = z(t) + \int_0^t z(t - y)dF(y) = z(t) + (F \odot Z)(t).$$

Notice here we can take $F$ as any monotone, non-decreasing function with $F|_{(-\infty,0)} = 0$ and $F(+\infty) < \infty$.

# 4 SIMPLE INVENTORY MANAGEMENT MODEL

## 4.1 Setup and parameters for the model

A storage facility has capacity $S$. Demand for storage within the $n^{th}$ month is a random variable $D_n$, where all $D_n$'s ($n \geq 1$) are iid and uniformly distributed on an integer interval. If at the end of a month the storage is empty, it is refilled with $S$ items immediately. Let $X_n$ denote the number of items in storage at the end of the $n^{th}$ month. Then

$$X_n := \begin{cases} (X_{n-1} - D_n)_+, & 0 < X_{n-1}; \\ (S - D_n), & X_{n-1} = 0. \end{cases}.$$

Note that $(X_n)_{n \geq 0}$ is a Markov chain with state space $\{0, 1, \dots, S\}$. The chain has a stationary distribution $\pi$.

## 4.2 Long-term average inventory levels

The average inventory level is

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i.$$

By the ergodic theorem, this limit exists and is equal to

$$\sum_{i=0}^{S} i \cdot \pi(i).$$

So to complete the calculation, we need to find the stationary distribution, which can of course be computed using the formula

$$\pi = (1, 1, \dots, 1)(I - P + \mathbb{1}_{(S+1)\times(S+1)})^{-1},$$

where $P$ is the transition matrix for the chain. We can compute the matrix $P$ explicitly using the definition of $X_n$ above.

## 4.3 Long-term average unsatisfied demand per month

The average sales per month is given by

$$\begin{aligned} \mathbb{E}[D_U] = \sum_{i=2}^{S} \pi(i)\mathbb{E}[D_n] &+ \pi(1)\mathbb{E}[\text{sales in n} - \text{th month}|1 \text{ item by its beginning}] \\ &+ \pi(0)\mathbb{E}[\text{sales in n} - \text{th month}|0 \text{ item by its beginning}]. \end{aligned}$$

Consequently, the average unsatisfied demand is $\mathbb{E}[D_n] - \mathbb{E}[D_U]$.

## 4.4 Average time between storage refills

Define $y_i$ as the time interval between storage refills. Set $S_0 = 0$, $S_n = \sum_{i=1}^{n} y_i$ for $n \geq 1$ as the associated renewal sequence. Then $S_n$ counts the number of storage refills with a counting function given by $N(t) := |\{n : S_n \leq t\}|$. By the LLN, almost everywhere we have that

$$\frac{N(t)}{t} \to \frac{1}{\mu} = \frac{1}{\mathbb{E}[y_i]}.$$

Thus $N(t)$ approximately denotes the total sales up to time $t$, divided by $|S|$.

# 5 MACHINE OPERATION PROBLEM

## 5.1 Preliminaries

A machine's operation interval has CDF $F$. After the machine breaks down, it takes $M$ dollars to repair. Alternatively, a basic check-up and replacement of worn-out details while the machine is still operational would cost $N < M$ dollars, and after that replacement, the machine continues to work as if it were new. The management wants to find the best moment of time after the start of an operation cycle to perform the check-ups and replacements. Find the optimal time by analysis of the corresponding renewal reward process.

## 5.2 Solution for uniform $F$

We will assume that $F$ is uniform on $[0, 12]$ (measured in months). Naturally, we will have to take $T \leq 12$. First, we define our corresponding renewal process: Let $y_i(T)$ denote the time interval between consecutive breakdowns or replacements. Then for $s_0 = 0$, $s_n(T) := \sum_{i=1}^{n} y_i(T)$ is the time of the $n^{th}$ replacement (or end of the $n^{th}$ operation interval). The reward $R_n(T)$ will denote the amount of money spent on the machine during this operation cycle. Note that $R_n(T)$ can take two values: $M$ if the machine breaks down before $T$, or $N$ is we get to do the basic replacement.

We will compute the distribution, then the expectation. We know that $y_i(T)$ takes values in $[0, T]$, and that conditioned on $y_i(T) < T$, it is uniform on $[0, T)$. So we have that

$$\mathbb{P}[U([0, 12]) \geq T] = 1 - \frac{T}{12}.$$

Thus,

$$\mu(T) := \mathbb{E}[y_i(T)] = \int_0^T \frac{t}{12} dt + T \cdot \mathbb{P}[y_i(T) = T] = T - \frac{T^2}{24}.$$

Now, we can compute that

$$\begin{aligned}
\mathbb{E}[R_i(T)] &= -M \cdot \mathbb{P}[y_i(T) < T] - N \cdot \mathbb{P}[y_i(T) = T] \\
&= -\frac{M}{12} \cdot T - N \left(1 - \frac{T}{12}\right) \\
&= -\frac{(M - N)}{12} \cdot T - N.
\end{aligned}$$

Finally, with $f(T) := \mathbb{E}[R_i(T)]/\mathbb{E}[y_i(T)]$, we can compute that

$$f'(T) = 0 \iff -\frac{(M - N)}{288} T^2 - \frac{N}{12} T + N = 0,$$

so by taking the quadratic solution inside the $T \in [0, 12]$ range, we get the optimal value for $T$: $T = \frac{12}{M-N} \left(\sqrt{N(2M - N)} - N\right)$.

## 5.3 Solution for exponential $F$

Assume $F \sim \text{Exp}(1)$. Again, $y_i(T)$ denotes the time between scheduled maintenences or brakdown, and have with $s_0 = 0$, $s_n(T) = \sum_{i=1}^{n} y_i(T)$. Now we can compute the expected reward as

$$\begin{aligned}
\mathbb{E}[R_i] &= -N \cdot \mathbb{P}[y_i(T) = T] - M \cdot \mathbb{P}[y_i(T) < T] \\
&= -Ne^{-T} - M(1 - e^{-T}).
\end{aligned}$$

By the same argument, this also implies that

$$\mathbb{E}[y_i(T)] = \int_0^T t \cdot e^{-t} dt + T \cdot \int_T^{\infty} e^{-t} dt$$

$$= -Te^{-T} + (1 - e^{-T}) + Te^{-T}.$$

Now we have that

$$\begin{aligned} f(T) &= \lim_{t\to\infty} \frac{R_i(t)}{t} = \frac{\mathbb{E}[R_i]}{\mathbb{E}[y_i]} \\ &= \frac{-Ne^{-T} - M(1 - e^{-T})}{1 - e^{-T}} \\ f(T) &= -M - \frac{N \cdot e^{-T}}{1 - e^{-T}}. \end{aligned}$$

In this case, we should let $T$ be as large as possible (and never perform checkups). Remember, to reconcile with our findings in this solution, the exponential distribution is memory-less.

## 5.4 Note on exam problem format

On the exam, we will be asked to work the problem with a concrete (simple) CDF $F$.

# 6 POISSON POINT PROCESSES (PPP)

## 6.1 Definitions

**Definition 6.1.** A *poisson point process* (PPP) on an interval $[a, b]$ can be defined as $\xi = \sum_{i=1}^{U} \delta_{t_i}$, where $U \sim$ Poisson($(b-a)\lambda$, the $t_i$ are uniform on $[a, b]$, and $U, t_1, t_2, \ldots$ are mutually independent. We also have the following constructions:

- A PPP on $[0, \infty)$ can be defined as the sum of mutually independent PPPs on $[n-1, n]$ for $n \geq 1$ as $\xi = \sum_{n \geq 1} \xi_n$ where $\xi_n$ is a PPP on $[n-1, n]$ and $\lambda$ is a paremeter called the *intensity* of the process.
- Given the above definition of PPP on $[0, \infty)$, the restriction of the process to any interval $[a, b]$ yields a PPP on $[a, b]$. Formally, we can write this for $\xi$ a PPP on $[0, \infty)$ as $\xi'(B) := \xi(B \cap [a, b])$ so that $\xi'$ is a PPP on $[a, b]$.

## 6.2 Connection to renewal sequences

Let $(S_n)_{n \geq 0}$ be a pure renewal sequence such that the interarrival distribution has the form $F(X) = 1 - e^{-\lambda X}$. Then the associated PP on $[0, \infty)$ given by $\xi = \sum_{n \geq 1} \delta_{S_n}$ is a PPP on $[0, \infty)$ with rate $\lambda$. Note that $|\{n \geq 1 : S_n \leq b\}| \sim$ Poisson($\lambda b$):

$$\mathbb{P}[|\{n \geq 1 : S_n \leq b\}| = m] = \mathbb{P}[S_m \leq b \wedge S_{m+1} > b] = \mathbb{P}[S_m \geq b] - \mathbb{P}[S_{m+1} \leq b].$$

## 6.3 Independence

**Proposition 6.2.** *Assume $\xi$ is a PPP on $[0, b]$ of rate (intensity) $\lambda$. Let $I_1, \ldots, I_k \subset [0, b]$ be disjoint intervals. Then $\{\xi(I_1), \ldots, \xi(I_k)\}$ are mutually independent Poisson random variables.*

## 6.4 Connection between the Poisson and normal distributions

**Proposition 6.3** (The Connection of Poisson and Normal Distributions)**.** *Let $\xi$ be a PPP on $[0, \infty)$ with rate $\lambda$. For each $t > 0$, denote by $F_t$ the CDF of $\frac{N(t)\mathbb{E}[N(t)]}{\sqrt{\mathrm{Var}(N(t))}} = \frac{N(t)-\lambda t}{\sqrt{\lambda t}}$, where $N(t) = \xi([0, t])$. Then as $t \to \infty$, $F_t \to \Phi$ in distribution a.e.*

## 6.5 Characterization of PPP

**Definition 6.4** (Stationary Increment Property)**.** A point process $\xi$ on $(0, \infty)$ is said to have the *stationary increment property* if for all $t' > t > 0$, we have $\xi((0, t']) - \xi((0, t]) \sim^d \xi((0, t' - t]).$

**Definition 6.5** (Independent Increment Property)**.** A point process $\xi$ on $(0, \infty)$ is said to have the *independent increment property* if for all $k$ and any $t_1 < t_2 < \cdots < t_k$, we have that the variables

$$\xi((0, t_1]), \xi((t_1, t_2]), \ldots, \xi((t_{k-1}, t_k]),$$

are mutually independent.

**Question 6.6.** *Are the independent and stationary increment properties sufficient to characterize the PPP?*

*Solution.* No. Consider $\mathbb{E}[X_i(s_i)]$ a pure renewal sequence with interarrival distribution $F(t) = 1 - e^{-t}$. Then $\xi := \sum_{i \geq 1} 2\delta_{s_i}$ satisfies the above properties, but is not a Poisson process. We have to add in additional properties to control the measure of infinitessemal intervals to characterize the PPP. □

**Proposition 6.7.** *Suppose that $\xi$ satisfies the stationary and independent increment properties. Suppose also that $\xi$ satisfies each of the following conditions for some fixed $\lambda > 0$:*

*(A) For all $\delta \in (0,1]$, $\mathbb{P}[\xi((0,\delta]) = 0] = 1 - \lambda\delta + o(\delta)$;*

*(B) For all $\delta \in (0,1]$, $\mathbb{P}[\xi((0,\delta]) = 1] = \lambda\delta + o(\delta)$.*

*Then $\xi$ is a PPP with parameter $\lambda$.*

*Proof (Sketch).* Note that (A) and (B) $\implies \mathbb{P}[\xi((0,\delta]) \geq 2] = o(\delta)$ for all $\delta \in (0,1]$. The key step is to compute the distrubution on $(0,t]$ with any $t > 0$ a fixed real number, say as $\mathbb{P}[\xi((0,t]) = k]$. Now, to do this, we split $(0,t]$ into $M$ subintervals: $(0, \frac{t}{M}], (\frac{t}{M}, \frac{2t}{M}], \ldots, (t - \frac{t}{M}, t]$. Then

$$\begin{aligned}
\mathbb{P}\left[\xi((\frac{(i-1)t}{M}, \frac{it}{M}]) \geq 2, \text{ for some } i \leq M\right] &\leq M \cdot \mathbb{P}\left[\xi((0, \frac{t}{M}]) \geq 2\right] \\
&= M \cdot o\left(\frac{t}{M}\right) \\
&= o(t) = o(1)
\end{aligned}$$

Now, we can see that

$$\begin{aligned}
\mathbb{P}[\xi((0,t]) = k] &= o(1) + \mathbb{P}\left[\xi((0,t]) = k \wedge \xi\left((\frac{(i-1)t}{M}, \frac{it}{M}]\right) \leq 1, \forall i \leq M\right] \\
&= o(1) + \binom{M}{k} \mathbb{P}\left[\xi\left((0, t/M]\right) = 1\right]^k \cdot \mathbb{P}\left[\xi\left((0, t/M]\right) = 0\right]^{M-k},
\end{aligned}$$

where the last equation follows because there is an $I \subset \{1, 2, \ldots, M\}$ of size $k$ such that $\xi\left(((i-1)\frac{t}{M}, \frac{it}{M}]\right) = 1$ for all $i \in I$ and is zero for all $i \notin I$. So the main calculation continues as

$$\mathbb{P}[\xi((0,t]) = k] = o(1) + \frac{M(M-1)\cdots(M-k+1)}{k!}\left(\frac{\lambda t}{M} + o(t/M)\right)^k \left(1 - \frac{\lambda t}{M} + o(t/M)\right)^{M-k}.$$

Letting $M \to \infty$, we see that

$$M(M-1)\cdots(M-k+1) \times \left(\frac{\lambda t}{M} + o\left(\frac{t}{M}\right)\right)^k \to (\lambda t)^k,$$

and that $\left(1 - \frac{\lambda t}{M} + o(t/M)\right)^{M-k} \to e^{-\lambda t}$, which in total corresponds to a Poisson$(\lambda t)$ random variable. $\square$

## 6.6 Other facts about the PPP

**Theorem 6.8** (Superposition)**.** *Let $(\xi_i)_{i\geq 1}$ be mutually independent PPPs on $(0,\infty)$ such that the intensity of $\xi_i$ is $\alpha_i$ where we assume that $\alpha := \sum_{i\geq 1} \alpha_i < \infty$. Then the process $\xi := \sum_{i\geq 1} \xi_i$ is a PPP of intensity $\alpha$.*

**Theorem 6.9** (Thinning)**.** *Let $\xi := \sum_{i\geq 1} \delta_{s_i}$ be a PPP on $(0,\infty)$. Suppose that $(b_i)_{i\geq 1}$ is a sequence of mutually independent random variables with $\xi$ having distribution* Bernoulli$(p)$*. Then the new process $\xi' := \sum_{i\geq 1} b_i \cdot \delta_{s_i}$ is a PPP of intensity $\alpha p$.*

## 6.7 Non-homogeneous Poisson processes

### 6.7.1 Construction

Fix a function $\alpha(t)$ which is strictly positive and continuous such that $\int_0^\infty \alpha(s)ds = \infty$. Define $m(t) := \int_0^t \alpha(s)ds$, which is strictly increasing with $m(+\infty) = \infty$. (By extension, we can take $m(A) := \{y \in (0,\infty) : \exists x \in A, y = m(x)\}$.) We define $m^{-1}(x) := \inf\{u : m(u) \geq x\}$. Finally, we define a point process on $(0,\infty)$ as

$$\xi' := \sum_{i\geq 1} \delta_{m^{-1}(s_i)},$$

where $(s_i)_{i\geq 1}$ is a renewal sequence with interarrival distribution given by $F(y) = 1 - e^{-y}$. Then $\xi'$ is called a *non-homogeneous PPP of intensity $\alpha(t)$*.

### 6.7.2 Properties

We have the following properties:

- If $A_1, \ldots, A_n \subset (0, \infty)$ are disjoint Borel subsets, then the variables $\xi'(A_1), \ldots, \xi'(A_n)$ are mutually independent;
- For any set $A \subset (0, \infty)$ such that $\int_{t \in A} \alpha(t)dt < \infty$, the variable $\xi'(A) \sim \text{Poisson}\left(\int_A \alpha(t)dt\right)$.

Proofs are given on page 56 of the notes.

## 6.8 Application: Model of periodic demand

**Example 6.10.** Let $\alpha(t) = 2 + \sin(t)$. Suppose that customers arrive according to PPP($\alpha(t)$). Then $\forall t_1 < t_2$, we have that the expected number of customers to arrive in the time interval $[t_1, t_2]$ is

$$\int_{t_1}^{t_2} \alpha(s)ds = 2(t_2 - t_1) - \cos(t_2) + \cos(t_1).$$

## 6.9 Example: Pure renewal sequences

For $S_0 = 0$, $(S_n)_{n \geq 0}$ is a pure renewal sequence with interarrival distribution $F(t) = 1 - e^{-\lambda t}$ for $t \geq 0$. The associated point process with this sequence on $[0, \infty)$ is

$$\xi = \sum_{i=1}^{\infty} \delta_{S_i},$$

which is a PPP with intensity $\lambda$. We define the counting function of this Poisson process to be $N(t) := |\{n : S_n \leq t\}|$. Therefore, $\mathbb{E}[N(t)] = 1 + \lambda t$.

# 7 CONVOLUTIONS OF FUNCTIONS

## 7.1 An atypical take on function convolutions

**Definition 7.1** (The Riemann-Stieltjes Integral)**.** Let $U(x)$ be a monotonic function, and $g$ be continuous on $\mathbb{R}$. For any increasing sequence $(x_k)_{k\in\mathbb{Z}}$, with $\lim_{k\to\infty} x_k = +\infty$ and $\lim_{k\to-\infty} x_k = -\infty$ and $\limsup_k (x_k - x_{k-1}) = 0$, we define $\int_{-\infty}^{\infty} g(x)dU(x)$ to be the limit

$$\sum_{k=-\infty}^{\infty} g(x_k)(U(x_k) - U(x_{k-1})).$$

If this latter limiting sum converges, then $\int g(x)dU(x)$ is called a *Riemmann-Stieltjes integral.* Note that if $U(x)$ is differentiable, then $\int gdU = \int g(x)U'(x)dx$.

**Definition 7.2** (Convolution and Its Key Properties)**.** We consider a monotone function $F$ with $F(\pm\infty) = \pm\infty$, either continuous on $\mathbb{R}$, or having only a finite number of discontinuities. Then for any locally bounded function $g$, we define the *convolution of functions* to be

$$(F \odot g)(t) := \int_{-\infty}^{\infty} g(t-x)dF(x), \forall t \in \mathbb{R}.$$

We have the following notable (key) properties of this type of function convolution:

- If $F$, $g$ are both identically zero whenever $t < 0$,

$$(F \odot g)(t) = \begin{cases} \int_{0^-}^{t} g(t-x)dF(x), & t \geq 0; \\ 0, & t < 0. \end{cases}$$

- (*Commutativity*): $F \odot G = G \odot F$;
- (*Associativity*): $(F \odot G) \odot H = F \odot (G \odot H)$.

Note that if $X, Y$ are independent random variables with respective CDFs, $F_X, F_Y$, then $F_{X+Y} = F_X + F_Y$. So we can also define the next notation for multiple convolution of CDFs: If $F$ is the CDF of a non-negative random variable $X$, and $(X_i)_{i\geq 1}$ are iid with the same distribution as $X$, then we denote the CDF of $X_1 + \cdots + X_n$ by $F^{\odot n}$. The special case of $F^{\odot 0} \equiv \mathbb{1}_{[0,\infty)}$ is the CDF of a random variable which is always zero. For all integers $n, m \geq 0$, we see that $F^{\odot n} \odot F^{\odot m} = F^{\odot(n+m)}$.

## 7.2 The renewal equation

**Definition 7.3** (Renewal Equation)**.** Suppose that $F$ is a strictly positive multiple of some CDF of a non-negative random variable, and that $Z(t)$ is a function on $\mathbb{R}$ such that $Z(t) = 0$, $\forall t < 0$. Then

$$\underline{Z}(t) = Z(t) + \int_0^{th} \underline{Z}(t-y)dF(y),$$

is called the *renewal equation* for an unknown function $\underline{Z}(t)$.

## 7.3 Solution of the renewal equation for pure renewal sequences

Consider a pure renewal sequence with iid $(Y_i)_{i\geq 1}$ defined by $s_n := \sum_{i=1}^{n} Y_i$ for $n \geq 1$. There is an associated counting function, $N(t) := |\{n \geq 0 : s_n \leq t\}|$. Assume $F$ is the *interarrival distrubution* (i.e., the CDF of the $Y_i$'s). Then for any $n \geq 1$, we have

$$\mathbb{P}[s_n \leq t] = \int_0^t \mathbb{P}[s_{n-1} \leq t - Y]dF(Y).$$

This is true because $s_{=}s_{n-1} + Y_n \implies \mathrm{cdf}(s_n) = F_{s_{n-1}} \odot F_Y$.

Let $U(t)$ be defined by

$$\begin{aligned} U(t) &= \sum_{n\geq 0} \mathbb{P}[s_n \leq t] = 1 + \sum_{n\geq 1} \mathbb{P}[s_n \leq t] \\ &= 1 + \int_0^t \sum_{n\geq 1} \mathbb{P}[s_{n-1} \leq t - Y] dF(Y), \end{aligned}$$

which implies that

$$U(t) = 1 + \int_0^t U(t-Y).$$

This $U$ satisfies the renewal equation with $Z = 1$, or more precisely, with $Z = \mathbb{1}_{[0,\infty)}$. (*The expectation of the counting function satisfies the renewal equation.*)

## 7.4 Solution of the renewal equation by convolutions

**Theorem 7.4.** *Let $U(t) := \sum_{n=0}^{\infty} F^{\odot n}(t) = \mathbb{E}[N(t)]$. Then $(U \odot Z)(t) = \int_0^t Z(t-y)dU(y)$ is a solution of the renewal equation.*

*Proof (Sketch).* By plugging in the solution, we obtain that

$$\begin{aligned} \underline{Z}(t) &= Z(t) + F \odot (U \odot Z) = Z + (F \odot U) \odot Z \\ &= Z + \left( \sum_{n\geq 1} F^{\odot n} \right) \odot Z \\ &= F^{\odot 0} \odot Z + \left( \sum_{n\geq 1} F^{\odot n} \right) \odot Z \\ &= U \odot Z. \end{aligned}$$

This checks out. So we conclude that the theorem is correct. □

# 8 AGE-DEPENDENT BRANCHING PROCESS

## 8.1 Definition of the problem at hand

Let $G$ represent the CDF of the lifetime of an individual (CDF of a non-negative random variable). After death, each individual leaves $k$ offspring with probability $p_k$, where $\sum_{i\geq 0} p_i = 1$. We assume that the process started at time zero by a simple individual of age zero. Define $(X(t))_{t\geq 0}$ to be the size of the population at time $t$ (a continuous time process). Further, define $\underline{Z}(t) := \mathbb{E}[X(t)]$. The goal below is to find a formula for $\underline{Z}(t)$.

## 8.2 Proof: The expectation formula satifies a renewal equation

### 8.2.1 Initial constructions

Let $L_1$ be the lifetime of the first individual (a random variable distributed $\sim G$). Let $N$ be the number of offspring produced in the first generation so that $\mathbb{P}[N = k] = p_k$. Let $m := \mathbb{E}[N] = \sum_{k\geq 0} k \cdot p_k$, and assume that $m < \infty$. Then we have each of the following observations:

- $\underline{Z}(t) = \mathbb{E}[X(t)] = \mathbb{E}[X(t)\mathbb{1}_{\{L_1\leq t\}}] + \mathbb{E}[X(t)\mathbb{1}_{\{L_1>t\}}$;
- If $L_1 > t$, then $X(t) = 1$, so the last term above is equal to $\mathbb{P}[L_1 > t]$;
- Let $(X_n)_{n\geq 1}$ be a sequence of mutually independent copies of $X$. Then conditioned on the event $\{L_1 \leq t\}$, we have that $X(t) = \sum_{i=1}^N X(t - L_1)$;
- Each of the $N$ offspring behaves just like the first individual, but with a clock starting at $L_1$, instead of zero.

### 8.2.2 Finishing the proof

Based on the observations made in the last subsection, we see that:

$$\underline{Z}(t) = \mathbb{P}[L_1 > t] + \mathbb{E}\left[\sum_{i=1}^N X(t - L_1)\mathbb{1}_{\{L_1\leq t\}}\right].$$

We compute that

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^N X(t - L_1)\mathbb{1}_{\{L_1\leq t\}}\right] &= \sum_{k\geq 0}\mathbb{E}\left[\sum_{i=1}^N X(t - L_1)\mathbb{1}_{\{L_1\leq t\}}\mathbb{1}_{\{N\leq k\}}\right] \\
&= \sum_{k\geq 0}\mathbb{E}\left[\sum_{i=1}^N X_i(t - L_1)\mathbb{1}_{\{L_1\leq t\wedge N\leq k\}}\right] \text{, since } N \text{ is independent from } L_1, X_i \\
&= \sum_{k\geq 0}\sum_{i=1}^N \mathbb{E}\left[X_i(t - L_1\mathbb{1}_{\{L_1\leq t\wedge N\leq k\}}\right] \\
&\geq \sum_{k\geq 0}\sum_{i=1}^N \mathbb{E}\left[X_i(t - L_1\mathbb{1}_{\{L_1\leq t\wedge N\leq k\}}|N = k\right] \cdot \mathbb{P}[N = k] \\
&= \sum_{k\geq 0}\sum_{i=1}^N \mathbb{E}[X_i(t - L_1)\mathbb{1}_{\{L_1\leq t\}}|N = k] \cdot p_k \\
&= \sum_{k\geq 0}\sum_{i=1}^N \mathbb{E}[X_i(t - L_1)\mathbb{1}_{\{L_1\leq t\}}] \cdot p_k
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{k\geq 0} p_k \times \sum_{i=1}^{k} \mathbb{E}[X_i(t-L_1)\mathbb{1}_{\{L_1\leq t\}}] \\
&= \left(\sum_{k\geq 0} k\cdot p_k\right) \times \mathbb{E}[X_i(t-L_1)\mathbb{1}_{\{L_1\leq t\}}] \\
&= m \times \mathbb{E}[X_i(t-L_1)\mathbb{1}_{\{L_1\leq t\}}].
\end{aligned}
$$

This implies that

$$
\underline{Z}(t) = \mathbb{P}[L_1 > t] + m\mathbb{E}_{L_1}\mathbb{E}_{X_1}[X_1(t-L_1)\mathbb{1}_{\{L_1\leq t\}}],
$$

the second term of which we know is expressable as

$$
m \times \int_0^t \mathbb{E}_{X_1}[X_1(t-y)]dG(y) = m \times \int_0^t \underline{Z}(t-y)dG(y).
$$

So we have that

$$
\begin{aligned}
\underline{Z}(t) &= \mathbb{P}[L_1 > t] + m\int_0^t \underline{Z}(t-y)dG(y) \\
&= 1 - G(t) + \int_0^t \underline{Z}(t-y)d(mG(y)),
\end{aligned}
$$

which is a renewal equation with $Z(t) := 1 - G(t)$.

Now we observe the general solution to renewal equations in convolutions outlined in Section 7.4 (on page 14) and apply its form in our case here. In particular, we have that in the branching process case, $U = \sum_{n\geq 0} m^n G^{\odot n}$, so that

$$
\begin{aligned}
\underline{Z}(t) &= 1 - G(t) + \int_0^t \underline{Z}(t-y)d(mG(y)) \\
&= \left(\sum_{n\geq 0} m^n G^{\odot n}\right) \odot \left((1-G(t))\mathbb{1}_{\{t\geq 0\}}\right).
\end{aligned}
$$

If $G$ is the CDF of a constant, so that $G(t) = 1 \iff t \leq T$, then $G^{\odot n} = \mathbb{1}_{[nT,\infty)}$, which implies that

$$
(1-G(t))\mathbb{1}_{\{t\geq 0\}} = \mathbb{1}_{[0,T]}.
$$

Moreover, we can see that

$$
\begin{aligned}
\underline{Z}(t) &= \sum_{n\geq 0} m^n \mathbb{1}_{[nT,\infty)} \odot \mathbb{1}_{[0,T]} \\
&= \sum_{n\geq 0} m^n \mathbb{1}_{[(n-1)T,nT]}.
\end{aligned}
$$

In summary, we obtain a piecewise constant function $\underline{Z}(t)$ which is exponentially increasing at regulat intervals.

## 8.3 Examples from the final review notes

Conditioned on the event that the first individual died within $[s, s+ds]$ leaving behind $k$ offspring, the expectation of $X$ at time $t > s$ is equal to $k$ times the (unconditional) expectation of $X$ as time $t - s$. In words, this describes the *self-similarity property*, which provides us with the following renewal equation formulation:

$$
Z(t) = 1 - G(t) + m \times \int_0^t Z(t-y)dG(y).
$$

# 9 TIME-HOMOGENEOUS CONTINUOUS MARKOV CHAINS

## 9.1 Definitions of these chains (on a countable state space)

**Definition 9.1** (Homogeneous Continuous-Time Markov Chain)**.** Let $S \subset \mathbb{Z}$ be a state space. The sequence $(X(t))_{t\in[0,\infty)}$ taking values in $S$ is a (homogeneous) *continuous-time Markov chain* if the following two properties are met:

- (*Markov property*): For all $m$, $\forall t_1 < t_2 < \cdots < t_m$:
$$\mathbb{P}[X(t_m) = j | X(t_1) = k_1, \ldots, X(t_{m-2}) = k_{m-2}, X(t_{m-1}) = i] = \mathbb{P}[X(t_m) = j | X(t_{m-1}) = i], \forall i, j, k_r \in S;$$
- (*Homogeneity*): For all $s, t > 0$,
$$\mathbb{P}[X(t+s) = j | X(t) = i] = \mathbb{P}[X(s) = j | X(0) = i].$$

## 9.2 Holding times

**Definition 9.2** (Holding Times)**.** For any state $i \in S$, we can consider the random variable $T(i)$, which denotes how much time we spend at state $i$ before moving on to a new state. We will assume that $\mathbb{P}[T(i) > 0] = 1$. This *holding time* is well defined because of the two properties defined above.

### 9.2.1 Explanation of why holding times are exponentially distributed

**Question 9.3.** *What is the distribution of $T(i)$?*

*Solution.* We condition on $X(0) = i$. Then $\forall t, s > 0$, we have that
$$\begin{aligned}
\mathbb{P}[T(i) > t+s | T(i) > s] &= \mathbb{P}[X(u) = i, \forall 0 \le u \le t+s | X(u) = i, \forall 0 \le u \le s] \\
&= \mathbb{P}[X(u) = i, \forall s < u \le t+s | X(u) = i, \forall 0 \le u \le s] \\
&= \mathbb{P}[X(u) = i, \forall s < u \le t+s | X(s) = i, \text{ by the Markov property} \\
&= \mathbb{P}[X(u) = i, \forall 0 < u \le t | X(0) = i], \text{ by homogeneity,} \\
&= \mathbb{P}[T(i) > t].
\end{aligned}$$

Thus if $F$ is the CDF of $T(i)$, we get that $\frac{1-F(t+s)}{1-F(s)} = 1 - F(t)$ for all $t, s > 0$. Define $g(u) := \log(1 - F(u))$. Then $g(t+s) = g(t) + g(s)$ for all $s, t > 0 \implies g$ is linear: $g(u) = -\lambda(i) \cdot u$ for some constant $\lambda(i) > 0$. This means that $T(i)$ is *exponential* with parameter $\lambda(i)$. ☐

## 9.3 Formal construction of $X(t)$

We have the following components to note in the formal construction of the $X(t)$:

- $(\lambda(i))$ are paremeters which describe the holding times;
- $Q$ is a transition matrix on $S$ with $Q_{ii} = 0$ for all $i \in S$;
- $(X_n)_{n\ge0}$ is the discrete Markov chain with transition matrix $Q$;
- $(E_n)_{n\ge1}$ is the collection of exponential random variables with parameter 1, all mutually independent with $(X_n)$.

For any realization of $(X_n)_{n\ge0}$ and $(E_n)$, we set $\widetilde{T}_n := \sum_{m=0}^{n-1} \frac{E_m}{\lambda(X_m)}$ for $n \ge 1$, with $\widetilde{T}_0 = 0$; and $X(t) = X_i$ for $\widetilde{T}_i \le t \le \widetilde{T}_{i+1}$ $(i = 0, 1, \ldots)$. We will suppose that $\sum_{m\ge0} \frac{1}{\lambda(X_m)} = +\infty$ almost surely, so that $\widetilde{T}_n \to \infty$ a.e. It can be checked that the process as defined satisfies the two conditions. The sequence $(X_n)_{n\ge0}$ is called the *embedded Markov chain.*

*Remark* 9.4. The most important point of this construction is that the holding times are independent from the transitions of the embedded Markov chain.
*Why must this be true?* Assume $X(0) = 1$. The probability that $X(t)$ jumps to $j$ after the initial state $i$, conditioned on the time initially spent at $i$ being at least $s$, is equal to:

$$\begin{aligned}
\mathbb{P}[\exists t_0 : X(t_0) = j \wedge X(t) = i, \forall 0 \leq t < t_0 | X(t) = i, \forall 0 \leq t \leq s] \\
&= \mathbb{P}[\exists t_0 : X(t_0) = j \wedge X(t) = i, \forall s < t < t_0 | X(t) = i, \forall 0 \leq t \leq s] \\
&= \mathbb{P}[\exists t_0 : X(t_0) = j \wedge X(t) = i, \forall s < t < t_0 | X(s) = i], \text{by Markov property} \\
&= \mathbb{P}[\exists t_0 > 0 : X(t_0) = j \wedge X(t) = i, \forall 0 < t < t_0], \text{by homgeneity,}
\end{aligned}$$

signaling a jump to $j$ after the initial state $i$.

From this, we see that $T(i)$ does not depend on $X(t)$ for $t \leq t_0$:

$$\mathbb{P}[T(i) > s] = \exp(-\lambda(i)s).$$

## 9.4 Discrete and continuous-time Markov chains (Homogeneous): A common viewpoint

**Definition 9.5.** Let $S$ be a finite state space and suppose that $(X_t)_{t\geq 0}$ is a discrete Markov chain. It is true that

$$\mathbb{P}[X_{t+s} = x | X_0 = x_0, \dots, X_t = x_t] = \mathbb{P}]X_s = x | X_0 = x_t], \forall s, t, x, x_0, \dots, x_t \in S,$$

*if and only if* there exists a stochastic matrix $P$ indexed over $S$ such that

$$\mathbb{P}[X_{t+s} = x | X_0 = x_0, \dots, X_t = x_t] = (P^s)_{x_t, x}.$$

This means that $(P^s)_{s\in\mathbb{N}}$ is a sequence of stochastic matrices associated with $(X_t)$.

**Definition 9.6.** Let $(Y_t)$ be a continuous-time Markov chain (homogeneous). Suppose that

$$\mathbb{P}[Y_{t+s} = y | Y_{t_1} = y_{t_1}, \dots, Y_{t_m} = y_{t_m}, Y_t = y_t] = \mathbb{P}[Y_s = y | Y_0 = y_t], \forall t, s \geq 0, \forall t_1 < \dots < t_m < t; \forall y, y_{t_1}, \dots, y_{t_m}, y_t \in S.$$

This condition is equivalent to a collection of stochastic matrices, $(P(t))_{t\geq 0}$ from an uncountable collection, such that

$$\mathbb{P}[Y_{t+s} = y | Y_{t_1} = y_{t_1}, \dots, Y_{t_m} = y_{t_m}, Y_t = y_t] = (P(s))_{y_t, y}.$$

Note that $\forall s_s, s_2 \geq 0$, $P(s_1) \cdot P(s_2) = P(s_1 + s_2)$.

*Remark* 9.7. We have the following remarks on the corresponding forms:

- We will have a generator matrix for the continuous case: $A = P'(0) = \lim_{\delta\to 0} \frac{P(\delta)-P(0)}{\delta - 0}$.
- In the discrete case, we have a version that looks like $A = P - I = \frac{P^1 - P^0}{1-0}$.
- Continuously, we have $P'(t) = AP(t) = P(t)A$. Whereas, discretely, we have something like $\frac{P^{t+1}-P^t}{(t+1)-t} = P^T \cdot A = A \cdot P^t$.

In the long term, we can make an analogy between the stationary distribution and mixing time of the discrete case and the $\lambda$'s of the continuous case (how long the process stays in each state parameters).

# 10 DISTRIBUTIONS OF TH MARKOV CHAINS

## 10.1 Forward and backward ODEs / PDEs for the matrix $P(t)$

Given a continuous-time Markov chain, $(X(t))$ with $(X_n)_{n\geq 0}$ an embedded Markov chain with transition matrix $Q$ and parameters $(\lambda(i))_{i\in S}$ which determine the holding times, we are interested in computing the distribution of $X(t)$.

**Definition 10.1** (Component Matrices)**.** For $i, j \in S$, define the *stochastic matrix* $P(t)$ by its components:

$$P_{i,j}(t) := -\mathbb{P}[X(t) = j | X(0) = i].$$

Define the *generator matrix* $A \equiv (A_{ij})_{i,j\in S}$ by

$$A_{ij} := \begin{cases} -\lambda(i), & i = j; \\ \lambda(i) \cdot Q_{ij}, & i \neq j. \end{cases}.$$

**Theorem 10.2.** *The matrix $P(t)$ is continuously differentiable as a function of $t$, and:*

- $P'(0) = A$;
- *(Backward equation):* $P'(t) = A \cdot P(t)$;
- *(Forward equation):* $P'(t) = P(t) \cdot A$.

*Remark* 10.3 (Solutions to the ODEs). By the theorem, we have that $p'(t) = p(t)A = Ap(t)$. So:

- If $p(t)$ is a real-valued function, and $A$ is constant, $p(t) = \exp(At)$.
- If $S$ is finite, then $p(t) = \exp(At)$.
- It helps to decompose the matrix-valued exponential function into its Taylor series expansion:

$$p(t) = \exp(At) = I + tA + \frac{t^2}{2!}A^2 + \frac{t^3}{3!}A^3 + \cdots.$$

*Remark* 10.4. Recall the definitions of the auxiliary variables, $(\widetilde{T}_n)$ from our construction of $X(t)$. We assume that $\widetilde{T}_n \to \infty$ a.e. as $n \to \infty$.

## 10.2 Proof sketch of the differential equation solution forms

### 10.2.1 Proof of the initial condition

*Proof.* Let $\Delta t$ be very small. Now, we denote by $(*)$

$$P_{ij}(\Delta t) = \mathbb{P}[X(\Delta t) = j | X(0) = i] \approx \mathbb{P}[\exists \widetilde{t} \in [0, \Delta t] : X(s) = i, \forall s \in [0, \widetilde{t}] \wedge X(s) = j, \forall s \in [\widetilde{t}, \Delta t] | X(0) = i],$$

since the probability of several jumps within the interval $[0, \Delta t]$ is negligible.
**Case 1:** $(i = j)$ Then $(*) = \mathbb{P}[\text{holding time at } i \text{ is at least } \Delta t] = e^{-\lambda(i)\Delta t} \approx 1 - \Delta t \cdot \lambda(i)$;
**Case 2:** $(i \neq j)$ Then

$$\begin{aligned} (*) &= \mathbb{P}[\text{holding time at state } i \text{ is at least } \Delta t | X(0) = i] \times \\ &\quad \times \mathbb{P}[\text{we jump from state } i \text{ to state } j \text{ before } \Delta t | \text{holding time at } i \text{ is less that } \Delta t] \\ &= \left(1 - e^{-\lambda(i)\Delta t}\right) Q_{ij} \\ &\approx \Delta t \cdot \lambda(i) Q_{ij}. \end{aligned}$$

**Next:** $P(0) = I$ (identity). Then $P(\Delta t) - P(0) \approx \Delta t \cdot A \implies P'(0) = A$. $\square$

### 10.2.2 Proof of the backward equation (Sketch)

We see that $\forall i, j \in S$,

$$\begin{aligned}
P_{ij}(t+\Delta t) &= \mathbb{P}[X(t+\Delta t) = j | X(0) = i] \\
&= \sum_{k \in S} \mathbb{P}[X(t+\Delta t) = j \wedge X(\Delta t) = k | X(0) = i] \\
&= \sum_{k \in S} \frac{\mathbb{P}[X(t+\Delta t) = j \wedge X(\Delta t) = k \wedge X(0) = i]}{\mathbb{P}[X(\Delta t) = k \wedge X(0) = i]} \cdot \frac{\mathbb{P}[X(\Delta t) = k \wedge X(0) = i]}{\mathbb{P}[X(0) = i]} \\
&= \sum_{k \in S} \mathbb{P}[X(t+\Delta t) = j | X(\Delta t) = k \wedge X(0) = i] \cdot \mathbb{P}[X(\Delta t) = k | X(0) = i].
\end{aligned}$$

Now by the Markov property, followed by the homogeneity property, we see that

$$\mathbb{P}[X(t+\Delta t) = j | X(\Delta t) = k \wedge X(0) = i] = d\mathbb{P}[X(t+\Delta t) = j | X(\Delta t) = k] = P_{kj}(t),$$

and we know that

$$\mathbb{P}[X(\Delta t) = k | X(0) = i] = P_{ik}(\Delta t).$$

Together, this implies that

$$\begin{aligned}
P_{ij}(t+\Delta t) &= \sum_{k \in S} P_{kj}(t) \cdot P_{ik}(\Delta t) & \implies \\
P(t+\Delta t) &= P(\Delta t) \cdot P(t) & \implies \\
P'(t) &= A \cdot P(t),
\end{aligned}$$

because $P(t+\Delta t) - P(t) = (P(\Delta t) - I)P(t)$.

### 10.2.3 Proof of the forward equation (Sketch)

We essentially repeat the same argument, using the identity

$$P_{ij}(t+\Delta t) = \mathbb{P}[X(t+\Delta t) = j | X(0) = i] = \sum_{k \in S} \mathbb{P}[X(t+\Delta t) = j \wedge X(t) = k | X(0) = i].$$

## 10.3 Other examples to consider

See the example from the 12/3/2019 lecture notes (page 65).

11

# ANALYSIS OF THE ON-OFF PROCESS

## 11.1 Description of the procedure

A machine has two states: 1 meaning 'on' (corresponds to holding time of parameter $\lambda_1$), and 0 meaning 'off' (corresponds to holding time with parameter $\lambda_0$). The machine is taken to be operational at the start: $X(0) = 1$. We take $X(t)$ to be the state of the machine at time $t$. The transition matrix is given by

$$Q = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

What is the distribution of $X(t)$?

## 11.2 Distribution of the process in time

The *generator matrix* for the problem is given by

$$A = \begin{bmatrix} -\lambda_0 & \lambda_0 \\ \lambda_1 & -\lambda_1 \end{bmatrix}.$$

Thus using the equation $P'(t) = P(t)A$, we see that

$$p_{00}'(t) = -\lambda_0 p_{00}'(t) + \lambda_1 p_{01}(t); p_{01}'(t) = \lambda_0 p_{00}(t) - \lambda_1 p_{01}(t).$$

So:

$$p_{00}' = -p_{01}'; p_{00}(0) = 1, p_{01}(0) = 0 \implies p_{01} = 1 - p_{00}.$$

So we have that $p_0'0 = \lambda_1 - (\lambda_0 + \lambda_1)p_{00}$ and the homogeneous solution $xe^{-(\lambda_0+\lambda_1)t}$ gives us that $p_{00}(t) = ce^{-(\lambda_0+\lambda_1)t} + \frac{\lambda_1}{\lambda_0+\lambda_1}$ where $p_{00}(0) = 1$. This shows that

$$p_{00}(t) = \frac{\lambda_0}{\lambda_0 + \lambda_1} e^{-(\lambda_0+\lambda_1)t} + \frac{\lambda_1}{\lambda_0 + \lambda_1}.$$

The other entries of $P(t)$ are obtained similarly.

## 11.3 Towards a general strategy for solution

We wish to decompose $A = B\Sigma B^{-1}$, where the columns of $B$ are eigenvectors of $A$, and $\Sigma$ is a diagonal matrix with eigenvalues of $A$ on its diagonal. Since

$$\det(A - SI) = (-\lambda_0 - s)(-\lambda_1 - s) - \lambda_0\lambda_1 = s(\lambda_0 + \lambda_1 + s),$$

it follows that the eigenvalues of $A$ are at $s = 0$ (with corresponding eigenvector $[1, 1]^T$) and $s = -(\lambda_0 + \lambda_1)$ (with corresponding eigenvector $[\lambda_0, -\lambda_1]^T$. Now we can see that

$$\begin{aligned} \sum_{n\geq 0} \frac{t^n}{n!} A^n &= B \begin{bmatrix} 1 & 0 \\ 0 & \sum_{n\geq 0} \frac{(-t)^n}{n!}(\lambda_0 + \lambda_1)^n \end{bmatrix} B^{-1} \\ &= B \begin{bmatrix} 1 & 0 \\ 0 & \exp(-(\lambda_0 + \lambda_1)t) \end{bmatrix} B^{-1}. \end{aligned}$$

Moving on, to invert $B$, we have that

$$B = \begin{bmatrix} 1 & \lambda_0 \\ 1 & -\lambda_1 \end{bmatrix} \implies B^{-1} = \frac{1}{\lambda_0 + \lambda_1} \begin{bmatrix} \lambda_1 & \lambda_0 \\ 1 & -1 \end{bmatrix}.$$

So we have that

$$\begin{aligned}
p(t) &= \begin{bmatrix} 1 & \lambda_0 \\ 1 & -\lambda_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \exp(-(\lambda_0+\lambda_1)t) \end{bmatrix} \frac{1}{\lambda_0+\lambda_1} \begin{bmatrix} \lambda_1 & \lambda_0 \\ 1 & -1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & \lambda_0 \exp(...) \\ 1 & -\lambda_1 \exp(...) \end{bmatrix} \frac{1}{\lambda_0+\lambda_1} \begin{bmatrix} \lambda_1 & \lambda_0 \\ 1 & -1 \end{bmatrix} \\
&= \frac{1}{\lambda_0+\lambda_1} \begin{bmatrix} \lambda_1 + \lambda_0 \exp(...) & \lambda_0 - \lambda_0 \exp(...) \\ \lambda_1 - \lambda_1 \exp(...) & \lambda_0 + \lambda_1 \exp(...) \end{bmatrix}.
\end{aligned}$$

We then can see that

$$\lim_{t\to\infty} p(t) = \frac{1}{\lambda_0+\lambda_1} \begin{bmatrix} \lambda_1 & \lambda_0 \\ \lambda_1 & \lambda_0 \end{bmatrix}.$$

*Remark* 11.1. If $S$ is finite with $|S| = n$, then $P(t)$ and $A$ are finite-dimensional matrices. Thus our general strategy is to find a diagonalization decomposition for $A$, and then go forward to represent $p(t)$ in the form

$$p(t) = B \begin{bmatrix} \exp(\xi_1 t) & & \\ & \ddots & \\ & & \exp(\xi_n t) \end{bmatrix} B^{-1}.$$

# 12 ANALYSIS OF THE LINEAR PURE BIRTH PROCESS

## 12.1 Defining the process type

Let $(X(t))$ be the process where $(X_n)_{n\geq 0}$ is the discrete embedded Markov chain over a discrete space $S = \{1, 2, 3, \dots\}$ with a transition matrix defined by $Q_{i,i+1} = 1$ and 0 otherwise (ones on the diagonal above the main diagonal). The *linear birth process* will be characterized by $\lambda(i) = \lambda \cdot i$ where $\lambda > 0$ is a parameter which determines the growth rate of the population. Here, $X(t)$ denotes the size of the population at time $t$ (starting with $X(0) = 1$). The linear growth of $\lambda(i)$ can be interpreted as follows: the rate of growth of the population is proportional to its current size.

### 12.1.1 Relation to age-dependent process

We have the following interpretation of the linear birth process as the age-dependent process: Let $G$ be the CDF of an exponential random variable with parameter $\lambda$ so that the lifetime of an individual, $G(t) = 1 - e^{-\lambda t}$. With $X(t)$ defined as before, we have that at the time of death of each individual, he/she leaves two descendants.

**Proposition 12.1** (The Exponential Distribution is Memoryless)**.** *If $\xi$ is distributed according to the CDF $G$, then $\mathbb{P}[\xi > t + s | \xi > t] = \mathbb{P}[\xi > s]$.*

**Claim 12.2.** *The age-dependent process $X(t)$ defined before coincides with the pure birth process also defined as before.*

*Proof Sketch.* Fix a time $t$ and condition on the event that the population size at time $t$ is equal to $i$. Let $\eta$ be the time interval before the population size increases to $i + 1$. *What is the distribution of $\eta$?*

$$\begin{aligned}
\mathbb{P}[\eta > s] &= \mathbb{P}[\text{at time } t + s, \text{ each of the } i \text{ individuals is still alive}] \\
&= \left(e^{-\lambda s}\right)^i = e^{-\lambda i s} \\
&\implies \eta \sim \operatorname{Exp}(\lambda(i)).
\end{aligned}$$

□

## 12.2 Derivation of the distribution using forward and backward time equations

For the *linear birth process*, we have that $A$ is a semi-diagonal matrix of the form

$$A = \begin{bmatrix} -\lambda & \lambda & & & \\ & -2\lambda & 2\lambda & & \\ & & -3\lambda & 3\lambda & \\ & & & \ddots & \ddots \end{bmatrix}.$$

Since the process can only increase the population, we have a requirement that for $j < i$, $p_{ij}(t) = 0$ (the lower triangle of $p(t)$ is all zeros). Now, we have that $p'(t) = p(t)A$, which implies that $p'_{11}(t) = (p(t) \cdot A)_{11}$, denoting the inner product of the first row in $p(t)$ and the first column of $A$, which is actually full of zeros for the most part from the $A$ matrix side. We thus obtain that

$$\sum_{i \geq 1} p_{1,i} \cdot A_{i,1} = p_{11} A_{11} = -\lambda p_{11}.$$

Since $p_{11}(0) = 1$ and $\mathbb{P}[X(0) = 1 | X(0) = 1] = 1$, we arrive at the solution

$$p_{11}(t) = c \exp(-\lambda t) = -\exp(-\lambda t).$$

Moving on, it is easy to see a generalization of this procedure. In particular, we have that $p'_{1n}(t) = (p(t) \cdot A)_{1n}$. We now 'hit' the non-zero terms of $A$ at the $(n-1)^{th}$ and $n^{th}$ rows, so $p'_{1n}(t) = (n-)\lambda p_{1,n-1}(t) - n\lambda p_{1n}(t)$. This is again a differential equation, however, since we do not know $p_{1,n-1}(t)$ we will need to inductively construct a solution. For $n := 2$, we get that $p_{12}(t) = \exp(-\lambda t)(1 - \exp(-\lambda t))$.

**Claim 12.3.** *We have that for all $n \geq 2$,*

$$p_{1n}(t) = e^{-\lambda t}(1 - \exp(-\lambda t))^{n-1}.$$

*Proof.* We have the initial conditions that $p_{1n}(0) = 0$. For $n \geq 2$, we have

$$\begin{aligned} p'_{1n}(t) &= -\lambda e^{-\lambda t}(1 - e^{-\lambda t})^{n-1} + (n-1)\lambda e^{-\lambda t}\left[1 - (1 - e^{-\lambda t})\right](1 - e^{-\lambda t})^{n-2} \\ &= -n\lambda p_{1n}(t) + (n-1)\lambda p_{1,n-1}(t). \end{aligned}$$

□

We finally arrive at the distribution:

$$\mathbb{P}[X(t) = n | X(0) = 1] = e^{-\lambda t}\left(1 - e^{-\lambda t}\right)^{n-1}.$$

To compute the expectation of $X(t)$, we use the PMF of the *geometric distribution*, which is $f(n) = p \cdot (1-p)^{n-1}$. In our case, we set $p := \exp(-\lambda t)$. So we have that

$$\begin{aligned} \mathbb{E}[X(t)] &= \sum_{n \geq 1} n \cdot p \cdot (1-p)^{n-1} \\ &= p \times \frac{1}{(1 - (1-p))^2} \\ &= \frac{1}{p} = e^{\lambda t}. \end{aligned}$$